1-1-2016

A Reliability Generalization Study of the Brief Symptom Inventory-18

Dareen Taha Alzahrani
University of Denver, dareen.alzahrani@du.edu

A RELIABILITY GENERALIZATION STUDY OF THE BRIEF SYMPTOM INVENTORY-18

_____

A Thesis

Presented to

the Faculty of the Morgridge College of Education

University of Denver

_____

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

_____

by

Dareen Taha Alzahrani

August 2016

Advisor: Dr. Antonio Olmos

Author: Dareen Taha Alzahrani
Title: A RELIABILITY GENERALIZATION STUDY OF THE BRIEF SYMPTOM INVENTORY-18
Advisor: Dr. Antonio Olmos
Degree Date: August 2016

Abstract

Reliability generalization (RG) is a meta-analytic method that aims to assess the variability of test score reliability across studies and identify the sources of this variability. In this study, a reliability generalization analysis was performed on studies of the Brief Symptom Inventory–18 (BSI-18) to examine the variability in Cronbach's alpha reliability estimates reported in the literature. This inventory was chosen because of its extensive use in counseling and medical settings and documented reliability and validity. The database that was consulted to collect articles was PsycInfo. The reported Cronbach's alphas were obtained to assess whether defined moderator variables affected reliability estimates. Out of the 161 references located, 48 studies met the selection criteria. For the Global Severity Index (GSI), the mean reliability was 0.91, 0.77 for the Somatic subscale, 0.85 for the Depression subscale, and 0.83 for the Anxiety subscale. The moderator analyses led to a predictive model where the type of population (clinical vs. nonclinical) for the GSI, and gender for the Somatic subscale were significant. Finally, clinical implications of the results are discussed.

.

Acknowledgements

I am grateful to the educators, friends, and family members who have supported this scholarly journey. I wish to acknowledge the Research Methods and Statistic department, particularly my thesis advisor, committee members, faculty and staff that have guided me throughout my master program.

I would first like to thank my thesis advisor Dr. Antonio Olmos. The door to Dr. Olmos office was always open whenever I had a question about my research. It has been a pleasure working with him during my master program and I appreciate all of his feedback and support.

I would also like to thank Dr. Kathy Green for her encouragement, professional insights, and support throughout my academic journey at University of Denver. I am especially thankful for her guidance during the writing of this report.

In addition, I am very thankful to my friends Priyalatha Govindasamy and Hanan Algamdi. I am gratefully indebted to them for their help in this thesis.

Finally, I must express my very profound gratitude to my lovely parents, sisters, brothers, and daughter and to my loving husband for providing me with unfailing support and continuous encouragement throughout my years of study. This accomplishment would not have been possible without them. Thank you.

Table of Contents

List of Figures

A Reliability Generalization Study of the Brief Symptom Inventory-18

Introduction

In psychometry, the concepts of reliability and validity are fundamental to the utility of any measure. Reliability can be defined as the consistency of scores on a test. This consistency can be estimated by different methods. According to classical test theory, score reliability is affected by different factors such as sample size and test length. A useful method to test the variability in score reliability estimates across a number of studies and to characterize the potential sources of this variance is reliability generalization (Vacha-Haase, Henson, & Caruso, 2002).

Meta-analysis is a quantitative method that is used to summarize and synthesize the results of several empirical studies. This technique is widely used in the field of medicine, psychology, and the social sciences (Hedges & Pigott, 2001). Reliability generalization (RG) is a meta-analysis technique that attempts to assess the variability of test score reliability across studies and identify the sources of this variability. Using the RG method helps researchers to identify the conditions under which the score reliability estimates of a particular test will be low, and the circumstances that will help to produce a more reliable score.

The Brief Symptom Inventory–18 is one of many assessments available to assess psychological distress. Being a screening tool that consists of only 18 items, the BSI-18 has an advantage over other assessments to measure psychological distress in that it is simple and easy to use. The applications of the BSI-18 include use in mental health contexts (Andjreu, et al., 2008) and medical settings (Merport & Recklitis, 2012). Despite the fact that the chosen inventory, the BSI-18, is widely applied in counseling and medical settings because of its simplicity and ease of use, no study has been conducted to provide a comprehensive evaluation of its score reliability across studies.

Literature Review

Reliability

As defined above, reliability is the consistency of scores on a test. This consistency can be estimated over time, forms, rater, and items. According to Mason (2007), reliability often is investigated by using a test–retest approach, which finds the correlation between the test score and the repeated administration of the same test. Reliability can also be estimated by finding the correlation between the test score and the score on a parallel form of the test. Another type of reliability is called internal consistency. "Internal consistency is concerned with the homogeneity of the items within a scale" (Devellis, 2012, p.34). This approach aims to "explore the degree to which random variation in test scores can be due to the consistency within the items" (Mason, 2007, p.30).

Vacha-Haase, Ness, Nilsson, and Reetz (1999) point out the importance of recognizing that the estimate of reliability is for the test score and not for the test. This corresponds with what Rowley (1976) states about the reliability of a test: "reliability refers to the score obtained by some sample of examinees on that test" (p.52). He explains that the measure itself is not reliable or unreliable; instead, the score on this

measure can be reliable or unreliable. Consistency depends on many factors such as the manner in which the measure was used, the group of examinees, and the conditions of the administration.

The importance of score reliability comes from Thompson's (1990) statement "measurement integrity is critical to the derivation of sound research conclusions" (p. 585). Vacha-Haase, Ness, Nilsson, and Reetz (1999) agree with Thompson's view that the reliability of scores affects the results of the data obtained from the measure and the interpretation of results.

Reliability Generalization

Haase (1998) believes that score reliability should be explored in all studies. She proposed the reliability generalization (RG) method that can be defined as "a measurement meta-analytic method used to explore the variability in score reliability estimates and to characterize the possible sources of this variance" (Vacha-Haase, Henson, & Caruso, 2002, p. 562). Reliability generalization is a meta-analysis that focuses on psychometric indices. Therefore, studies and not individuals are the units of the analysis and comprise the sample for an RG study.

Haase (1998) describes RG as a powerful method that can be used to identify the source of variance in the reliability estimate. RG studies aim to find characteristics that can predict the variability in a reliability estimate for a specific measure. Warne (2008) points out that RG is the most useful tool to substantiate that the reliability is a property of the scores on a test and not the test. RG studies show that the score reliability of a

specific test may systematically differ from study to study depending on some characteristics called moderators.

Vacha-Haase, Henson, and Caruso (2002) stress that the results of RG studies provide valuable information that can be used to improve the theoretical understanding of reliability. In addition, they mention that RG studies increase awareness about the sample characteristics that might affect the reliability of a score on a test. RG is a useful tool for test administrators and researchers to gain a better understanding of using a test and making decisions based on the results of the test.

Such claims imply that RG methods can indicate which sample and study features can affect the score reliability estimate of a given test. This provides an important implication of the circumstances that may yield a high estimate of a test score reliability and the situations that need to be avoided because a low estimate of score reliability is generated from them.

Reviewing the RG literature provides evidence of the value of this method. In López-Pina et al.'s (2015) RG study of the Yale–Brown Obsessive Compulsive Scale, researchers found that the standard deviation of the total test and the target population (clinical vs. nonclinical) could be used as predictor variables; these two variables explained 38.6% of the variability in coefficient alpha. Sun and Wang's (2015) RG study of the Children's Depression Inventory found that the length of the test affected the reliability; the score reliability was higher in the long form of the test. Also, researchers

concluded that the different language forms of the test did not affect the reliability which indicates the cross-cultural equivalence of score reliability.

The Brief Symptom Inventory–18

One of the inventories that is worthy of investigation using RG is the Brief Symptom Inventory-18. The Brief Symptom Inventory -18 (BSI-18) was chosen because of its extensive use in counseling and medical settings and documented reliability and validity as a measure of symptoms related to mental health. The BSI-18 is the most recent and short form of a series of instruments that were designed by Derogatis in 2011. Derogatis developed the Symptom Checklist-90 that consists of 90 items distributed over nine subscales. He then developed a short form of this checklist; this form, "The Brief Symptom Inventory," comprises 53 items and nine subscales (Merport & Recklitis, 2012). The BSI-18 was then developed to improve the structural validity of the BSI-53. According to Meijer, de Vries, and van Bruggen (2011), the results of many studies indicated that the BSI-18 can be described as unidimensional. Derogatis points out that

> The structural validity has improved [with the BSI-18] because the reduced scale is composed of only three dimensions—namely, somatization, depression, and anxiety—which together are more homogeneous than other dimensions from previous instruments, both conceptually and empirically (as cited in Meijer, de Vries, & van Bruggen, 2011, p. 193).

The Brief Symptom Inventory–18 is a self-report symptom checklist that consists of 18 items distributed over three subscales: Somatization, Anxiety, and Depression.

Screening for distress in clinical practice is an important issue in the field of psychology and psychiatry. With the advantage provided by its simplicity and ease of

application, the BSI-18 has been widely used to identify psychological distress in cancer survivors (Merport & Recklitis, 2012), patients with psychiatric disorders (Andjreu et al., 2008), patients with temporomandibular disorders (Durá et al., 2006), survivors of traumatic brain injuries (Lukow et al., 2015), patients with voice concerns (Misono et al., 2014), and also studies of drug users((Wang, Kelly, Liu, Zhang, & Hao, 2013). The Brief Symptom Inventory-18 has been translated and adapted in several languages: Spanish (Asner-Self, Schreiber, & Marotta, 2006), Chinese (Wang et al., 2013), Hebrew (Slone & Mayer, 2015), and German (Spitzer et al., 2011).

Despite acceptable psychometric properties frequently reported for the BSI-18 in published research studies, no study has carried out a comprehensive evaluation of its score reliability across studies. The present study fills this gap by meta-analyzing score reliability estimates obtained from a number of research studies. Results of this study will help researchers or practitioners understand the use of the BSI-18. In other words, RG gives information about the population and the sample characteristics that are appropriate to administer the instrument to, so taking these factors into account will ensure more knowledgeable estimation of the reliability of using the BSI-18 and greater understanding in its application.

Research Questions

The research questions were:

1. What is the average reliability for the BSI-18 across studies?
2. What is the average reliability for each subscale of the BSI-18 across studies?

3. What factors are associated with observed variance in BSI-18 reliability estimates?

Method

Sample of Published Studies

Previous studies using any of the three subscales of the BSI-18 were identified through an electronic search of the PsychInfo database using the keyword *Brief Symptom Inventory -18, Brief Symptom Inventory -18 AND Reliability, Brief Symptom Inventory -18 AND Cronbach's, BSI -18, BSI -18 AND Reliability, BSI -18 AND Cronbach's*. Initial search results produced 246 studies that used the BSI-18. The researcher imposed a limiter to identify studies published between 2001 and 2016, yielding 242 results. This time limiter was chosen depending on the inventory published year. After removing duplicated studies and studies published in languages other than English, the final sample comprised 161 studies. For all of the 161 selected articles, the full article was obtained and reviewed to assess the fit with the inclusion criteria. Studies meeting the following criteria were selected: (a) an empirical study where the BSI-18 was applied to the sample, (b) reported Cronbach's alpha with data from the study sample, and (c) was written in English.

Of the 161 articles reviewed, 21% failed to mention reliability, 19% reported alpha coefficients from another source, 7% provided separate alpha coefficients (alpha for more than one sample) or a range of alpha coefficients, 7% were not independent of some included studies, 5% were not about the BSI-18, 4% were books or articles that

researcher could not access, 2% provided another type of reliability, and 2% did not provide useful descriptive information. Forty-eight articles remained that included Cronbach's alpha reported from the study sample with sufficient descriptive information. Of the 48 articles, 15 unpublished dissertations were included.

Coding Procedure

To examine potential relationships between the reliability estimates and the study features, both the Cronbach's alpha and possible moderator variables related to the instrument and the study participants were coded. These coded study characteristics were selected based on a review of the RG literature. In a review study of RG studies that was conducted by Vacha-Haase and Thompson (2011), the results showed that using the number of items, score standard deviation, gender, and participant's age as predictors were among the better predictors of variability in score reliabilities. In another review and evaluation of RG studies, Henchy (2013) found that the majority of RG studies coded the sample size, gender, and participant's age as the sample characteristic that might influence the coefficient alpha. Thus, the following characteristics were coded: (a) sample size, (b) female percent, (c) mean age of participants, (d) type of sample, and (e) language. Also, the publication year was coded to examine change in score reliabilities over time, and the research quality (published vs. unpublished) was coded to test publication bias.

Inter rater reliability

According to Dieckmann, Malle, and Bodner (2009), "unreliability in the coding procedures adds additional random variation to the analysis, weakening the reliability and power of the results. At a very basic level, this can be addressed by employing multiple coders and assessing inter-rater reliability" (p. 103). Thus, in order to examine the reliability of the coding process, a second qualified coder coded eight articles (16% of the sample). The researcher created a coding sheet of the relevant variables to be used when coding studies. See Appendix A for the complete codebook, and Appendix B contains the coding sheet. The inter-rater reliability was calculated by the percent agreement method. Initially, raters had an agreement rate of 96% and after issues were resolved, raters reached 100% agreement.

Analysis

Cronbach's alphas were meta-analyzed in two steps: transformation and weighting. Coefficient alphas were transformed by means of the Hakstian and Whalen transformation formula in order to normalize the distribution of alpha which is usually skewed.

$$ES_{\alpha} = \sqrt[3]{1-\alpha} \quad (1)$$

Where *ES* is the effect size, and α is the coefficient alpha. Even though the Fisher's Z transformation is commonly used, the Hakstian-Whalen transformation is recommended by Rodriguez and Maeda (2006) because it is noted that Fisher's Z introduces bias in reliability generalization studies. The reliability coefficients were weighted by the inverse variance using the following formula

$$w = \frac{1}{\tau^2 + SE^2} \quad (2)$$

The symbol τ^2 represents the between study variance, and *SE* is the standard error of the effect size.

The heterogeneity exhibited by the reliability estimates was assessed with the Q statistic. Finally, moderator analyses were conducted through regression analyses assuming mixed-effects model. For conducting mixed effect model analyses, Card (2012) points out that this model is useful for evaluating some moderators and to generalize the

results beyond the studies included in the meta-analysis. This correspond with Rodriguez and Maeda's (2006) recommendation for RG authors to use a random effects model or mixed effects model to generalize their inferences beyond the studies included in the meta-analysis. Thus, this model was applied for three reasons: using sample and inventory characteristics as moderators, assuming that the reliability coefficient estimates came from different populations, generalizing the results beyond the included studies.

To facilitate interpretation of results, the average reliability estimates, and their confidence limits were back-transformed to the original metric of reliability coefficients by using the following formula:

$$a = 1 - ES_a^3 \qquad (3)$$

Results

Mean Reliability and Heterogeneity

Table 1 shows the main summary statistics for coefficient alpha. Even though the included studies were 48 articles, not every study reported the Cronbach's alpha for each subscale or for the total score. Therefore, the number of studies for each subscale is different. The 44 estimates reported for the total scale GSI yielded a (weighted) mean coefficient alpha of 0.91 (95% confidence limits: 0.89 and 0.92). For the Somatic subscale, coefficient alpha was computed from 29 different samples, leading to an overall estimate of 0.77 (confidence limits: 0.74 and 0.80. Thirty two estimates reported for the Depression subscale yielded a mean coefficient alpha of 0.85 (95% confidence limits: 0.84 and 0.87). An average coefficient of 0.83 (limits: 0.81 and 0.85) was found for the Anxiety subscale. Table 1 also presents the results of the *Q* statistics for the assessment of the variability exhibited by the reliability estimates. Coefficient alpha for the total scale and subscales showed statistically significant heterogeneity. Therefore, analyses to explain part of that heterogeneity were conducted. The results of the different studies, with 95% CI for GSI, Somatic, Depression, and Anxiety are shown in Figures 1, 2, 3, and 4 respectively.

Table 1

Summary Statistics for Coefficient Alpha

Scale	K	M	CI	Q
GSI	44	0.91	{0.89, 0.92}	561.46***
Somatic	29	0.77	{ 0.74 ,0.8}	411.81***
Depression	32	0.85	{0.84, 0.87}	248.77***
Anxiety	28	0.83	{0.81,0.85}	283.93***

Note. K = number of studies, M = mean Cronbach's alpha, CI = 95% confidence interval, Q = Hedge's Q

*** p<.0001

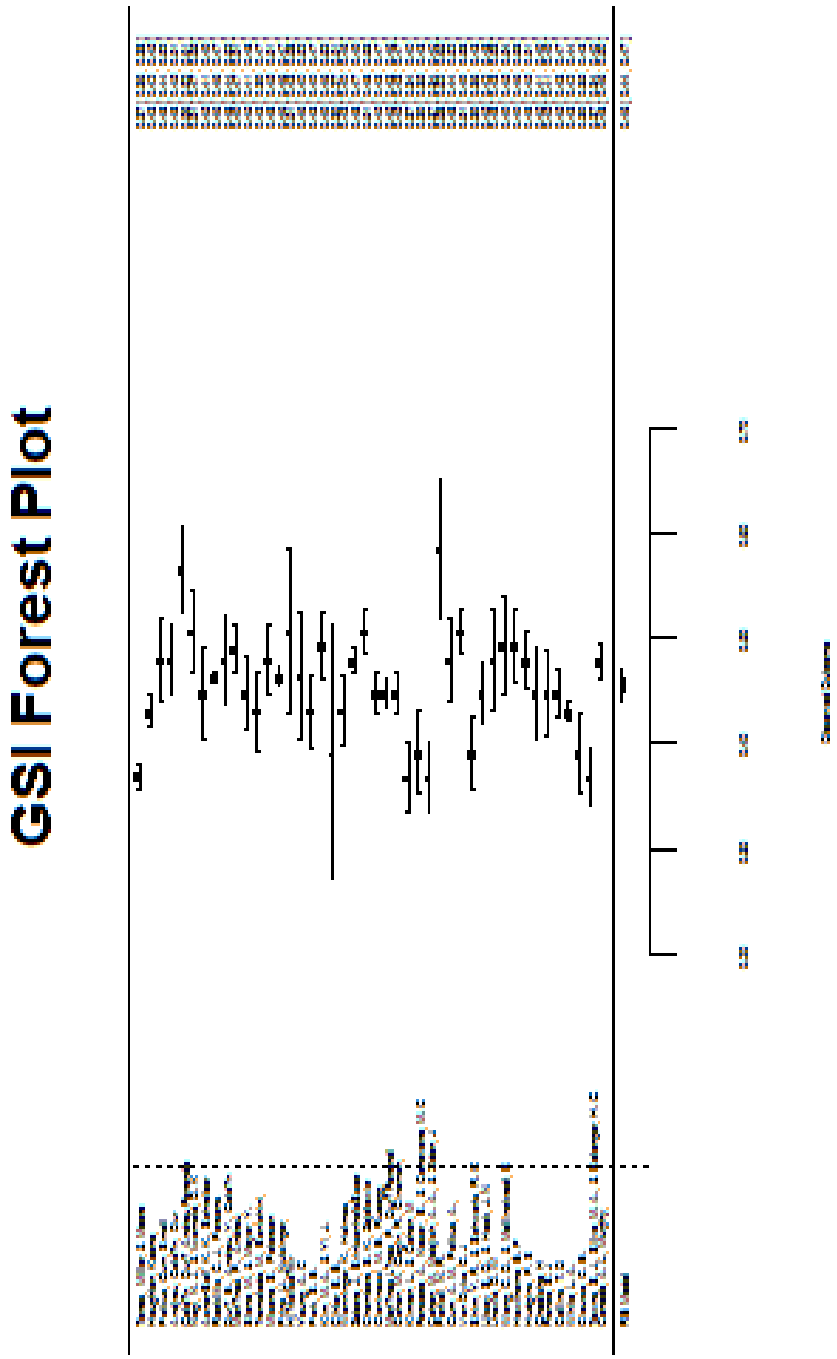

Figure 1. Forest Plot of the GSI

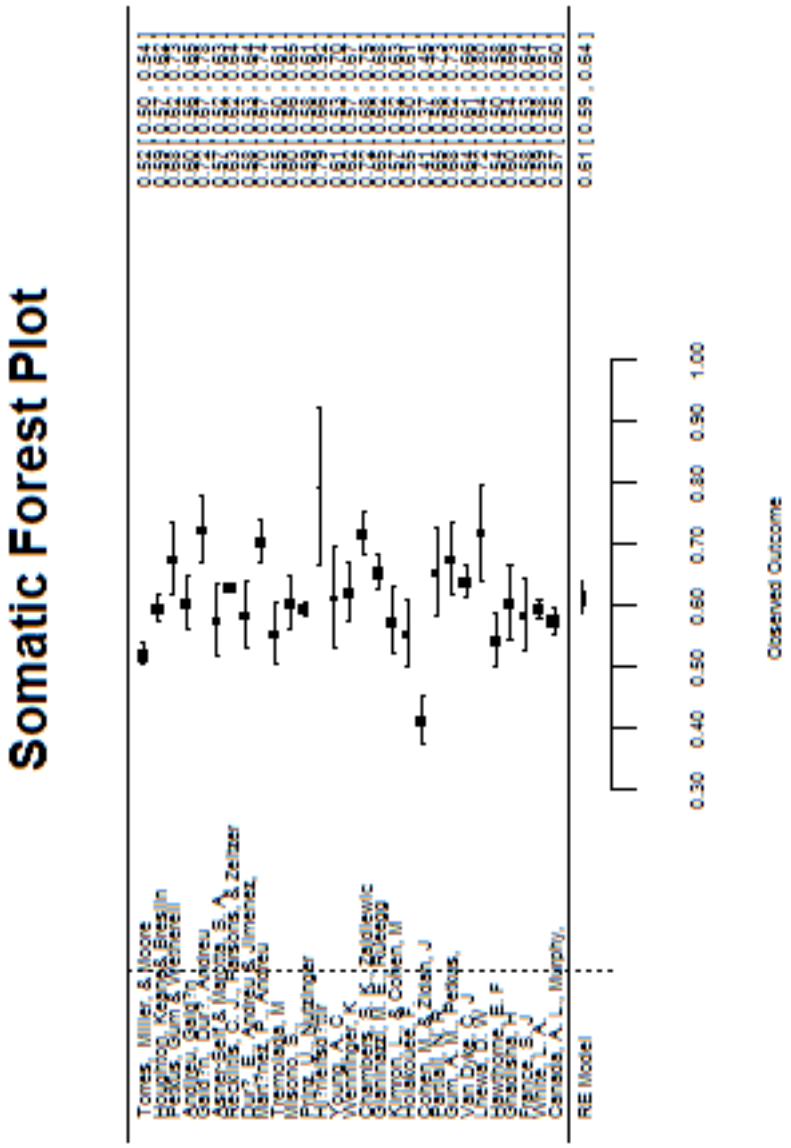

Figure 2. Forest Plot of the Somatic Subscale

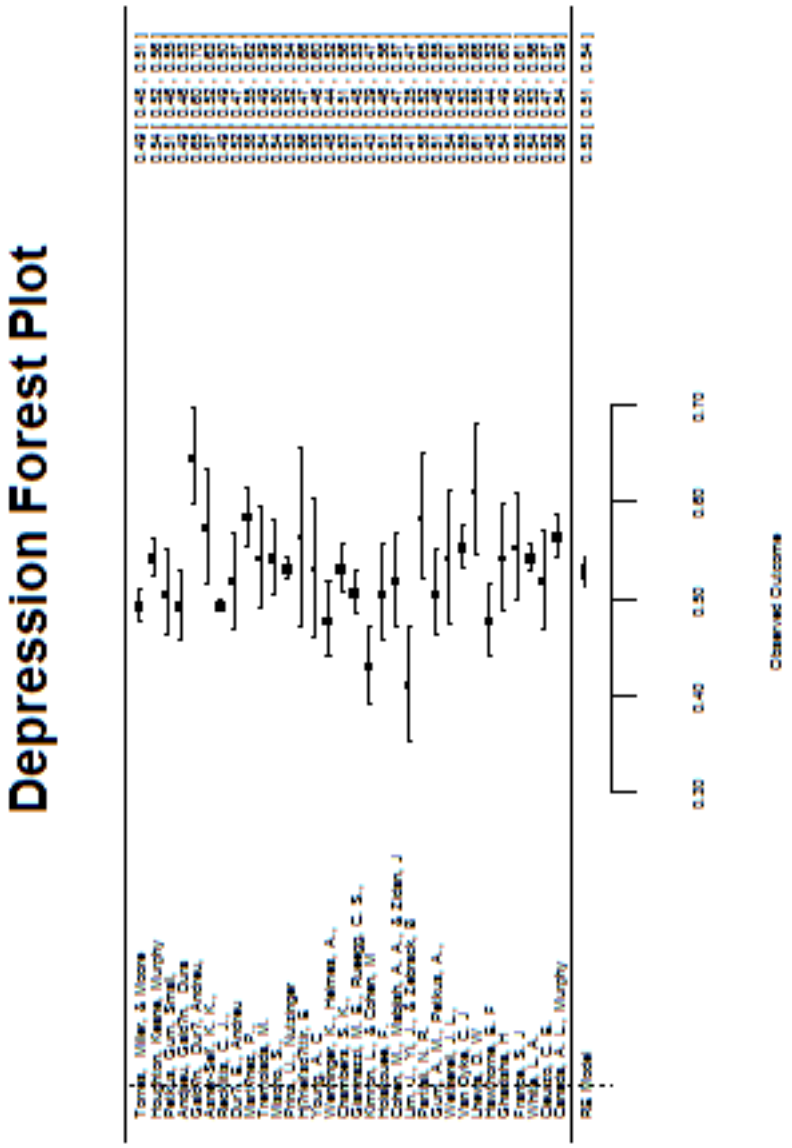

Figure 3. Forest Plot of the Depression Subscale

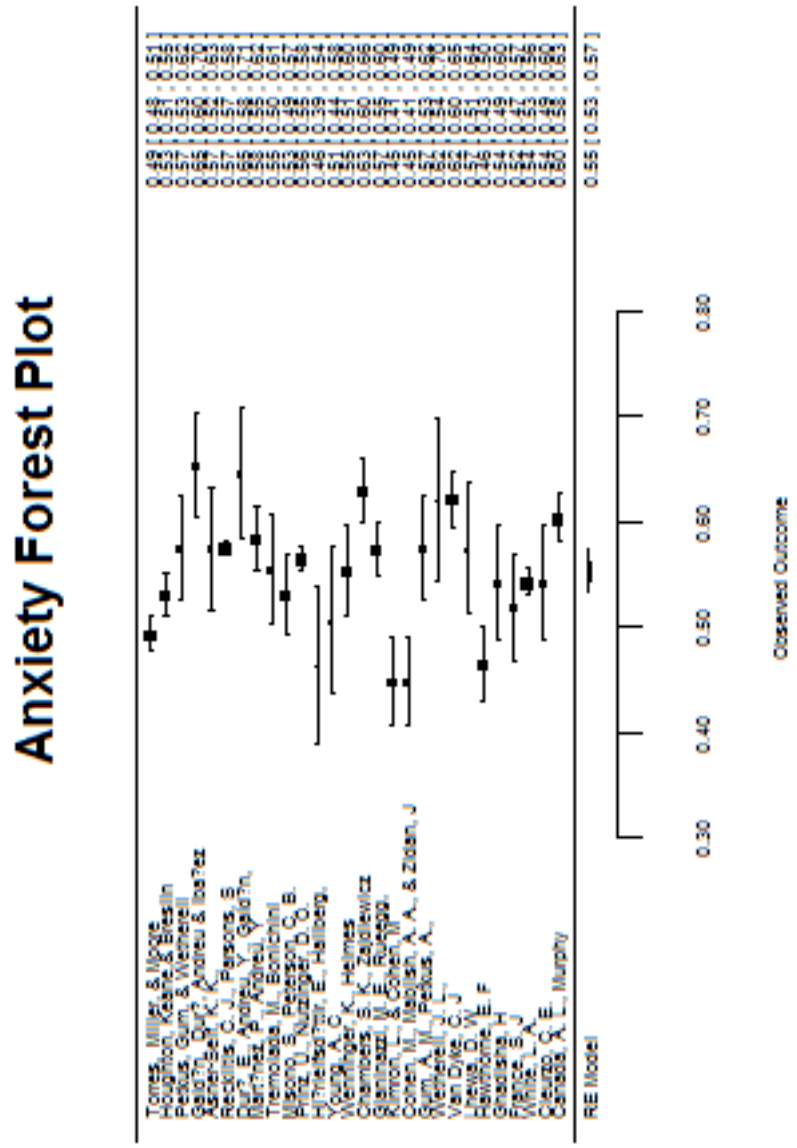

Figure 4. Forest Plot of the Anxiety Subscale

Moderator Analyses

A multiple regression analysis was used to assess whether year of publication, sample size, gender, mean age, population type, and language could predict the BSI-18 reliability scores. The continuous variables in the regression analysis were publication year, sample size, mean age, and gender (percentage of females in a study). The categorical variables in the regression analysis were population type (clinical / non clinical/ both clinical and non clinical), and language (English/ non English/ English and another language). More detailed results for each subscale are provided below.

GSI

Forty four sample coefficient alpha estimates were reported for the GSI scale. Six studies were eliminated from the multiple regression analysis because values for female, population type, and mean age were missing. All together, the six variables did not produce significant results in predicting the reliability estimate. A series of the hierarchical multiple regression analyses were conducted using different moderators in each model. The results did not produce significant increments to R^2. However, running simple regression analysis for the dummy coded variable (population type) shows that this moderator accounted for 24.85% of the variation in reliability with a significant result for the clinical sample ($b = 0.46$, $p < .0001$) and for the sample that had both clinical and non-clinical population ($b = -0.07$, $p = .0006$).

Somatic subscale

Twenty nine sample coefficient alpha estimates were reported for the Somatic scale. Two studies were eliminated from the multiple regression analysis because population type, and mean age have missing values. All together, the six variables did not produce significant results in predicting the reliability estimate. It was conducted a series of the hierarchical multiple regression analyses using different moderators in each model. The results did not produce significant increments to R^2. However, running simple regression analysis for gender shows that this moderator was able to account for 18.80% of the variation in reliability score (R^2=0.19, b = -0.0012, p = 0.0136).

Depression subscale

Thirty two sample coefficient alpha estimates were reported for the Depression scale. Two studies were eliminated from the multiple regression analysis because population type has missing values. All together, the six moderators did not produce significant results in predicting the reliability estimate. It was conducted a series of the hierarchical multiple regression analyses using different moderators in each model. The results did not produce significant increments to R^2. For further examination, a simple regression analysis was conducted to test population type moderator; it was able to account for 11% of the variation in reliability score but the results was not significant.

Anxiety subscale

Twenty eight sample coefficient alpha estimates were reported for the Anxiety scale. Three studies were eliminated from the multiple regression analysis because

population type, and mean age have missing values. All together, the six variables did not produce significant results in predicting the reliability estimate. It was conducted a series of hierarchical multiple regression analyses using different moderators in each model. The results did not produce significant increments to R^2.

Publication bias

Publication bias, which is also called the file drawer problem, is considered one of the threats that affect the validity of a meta-analysis. According to Dalton, Aguinis, Dalton, Bosco, and Pierce (2012),

> The file drawer problem rests on the assumption that statistically non-significant results are less likely to be published in primary level studies and less likely to be included in Meta analytic reviews, thereby resulting in upwardly biased Meta analytically derived effect sizes (p. 221).

In order to avoid this bias, the researcher included unpublished dissertations (K = 15 of the sample). In addition, the risk of publication bias was assessed by using the funnel plot method. This technique is a visual way to evaluate publication bias in meta-analysis that was introduced by Light and Pillemer (1984). In this graphic, standard error is on the y-axis and effect size is on the x-axis, and a dot represents each study. If there is a publication bias, the funnel plot will look asymmetrical. Figures 5 through 8 show the funnel plot using standard error on the y-axis. From Figures 5 and 7, it is clear there was publication bias in GSI and Depression, because there is a lack of balance between the two sides of the plot.

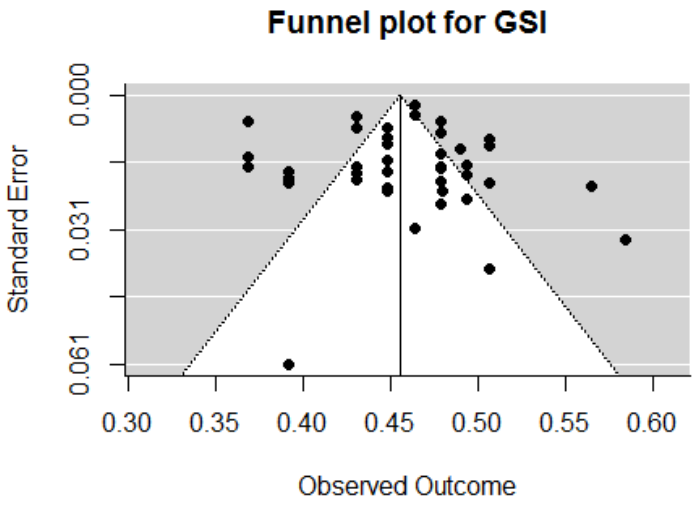

Figure 5. Funnel Plot for GSI

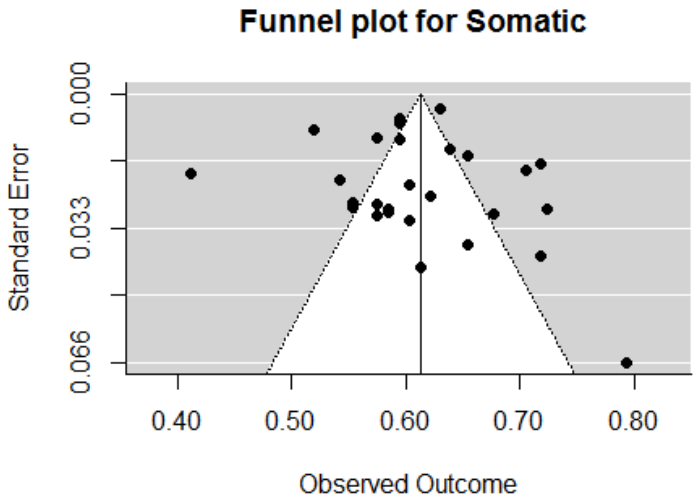

Figure 6. Funnel Plot for Somatic

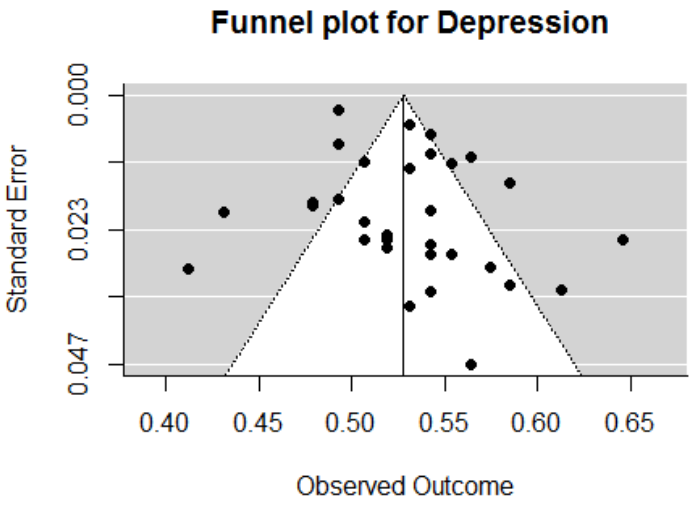

Figure 7. Funnel Plot for Depression

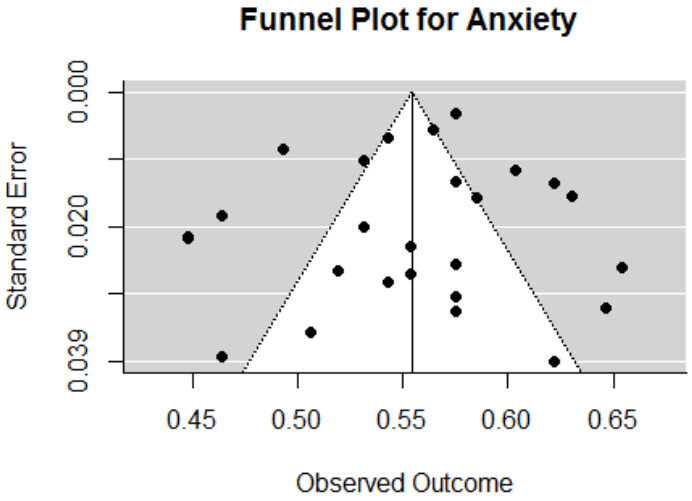

Figure 8. Funnel Plot for Anxiety

In order to reduce the subjectivity of evaluating the funnel plot, moderator analysis, regression, Kendall's rank correlation, the Egger's linear regression test, and the trim and fill method were used. Using multiple approaches to deal with the file drawer problem helps the researcher to determine the number of unpublished studies that need to be added to affect the effect size estimate.

Moderator analysis is recommended by Card (2012) as "one of the best methods to evaluate the potential impact of publication bias is to include unpublished studies in the meta-analysis and empirically evaluate whether these studies yield smaller effect sizes than published studies" (p. 262). Moderator analyses indicated a nonsignificant difference between published and unpublished studies (b = -0.009) for GSI, (b= 0.012) for Somatic, (b= -0.023) for Depression, and (b= -0.009) for Anxiety with $p > 0.05$ for all analyses.

Regression approaches are used to evaluate the funnel plot asymmetry and this approach has advantages over visual inspection of funnel plots because it reduces subjectivity by providing results that can be evaluated in term of statistical significance. The absence of statistically significant results indicates the absence of publication bias (Card, 2012, p. 267). The current study examined symmetry by regressing effect sizes on sample sizes. For GSI and all subscales, the results indicate the absence of an association between effect size and sample size because they were not statistically significant, $F(1,42)= 0.017$ for GSI, $F(1,27)= 0.002$ for Somatic, $F(1,30)= 0.443$ for Depression, and $F(1,26)= 0.2127$ for Anxiety; all had $p > 0.05$. Therefore, these results suggest the absence of publication bias.

Kendall's rank correlation was used to examine the correlation between the effect size and the standard error, and if it is not significant, that means there is absence of publication bias (Card, 2012, p. 266). The results for GSI and all subscales were nonsignificant. The Kendall's tau values were 0.0444 for GSI, 0.2315 for Somatic, 0.0484 for Depression, and -0.0370 for Anxiety; all had $p > 0.05$. Thus, these results indicate the absence of publication bias.

The Egger's test formally evaluates asymmetry of funnel plots by regressing the standard normal deviate of the effect size of each study from zero on the study precision. The possibility of publication bias can be indicated by a significant intercept (Card, 2012). The results for GSI and all subscales were not statistically significant, Z= 0.99 for GSI, 1.91 for Somatic, 0.96 for Depression, and -0.10 for Anxiety; all of them had $p > 0.05$. These results indicate the absence of publication bias.

The Trim and Fill approach is used to correct publication bias and involves a two-step iterative procedure to provide more accurate estimates of both mean effect size and the heterogeneity around this effect size (Card, 2012, p. 273). The trim step involves temporarily removing studies until a symmetric funnel plot is obtained then estimating an unbiased mean effect size for the remaining studies in the second step. In contrast, the Fill step reinstates the previously trimmed studies and then imputes studies in the underrepresented section until obtaining a symmetric funnel plot (Card, 2012, p. 273-274). The results of Trim and Fill approach for GSI indicated there were three missing studies on the left side needed to correct the effect size. However, the corrected effect size was 0.91, which is the same as the uncorrected effect size (0.91). See Figure 9. The

results of the Trim and Fill approach for Somatic indicated there were no missing studies needed to correct the effect size and the corrected effect size was 0.77, which was the same as the uncorrected effect size (0.77). See Figure 10. For Depression, the result indicated there were seven missing studies on the left side needed to correct the effect size and the corrected effect size was 0.86, which was greater than the uncorrected effect size (0.85). See Figure 11. For Anxiety, the result indicated that there were no missing studies needed to correct the effect size and the corrected effect size was 0.83, which is the same as the uncorrected effect size (0.83). See Figure 12.

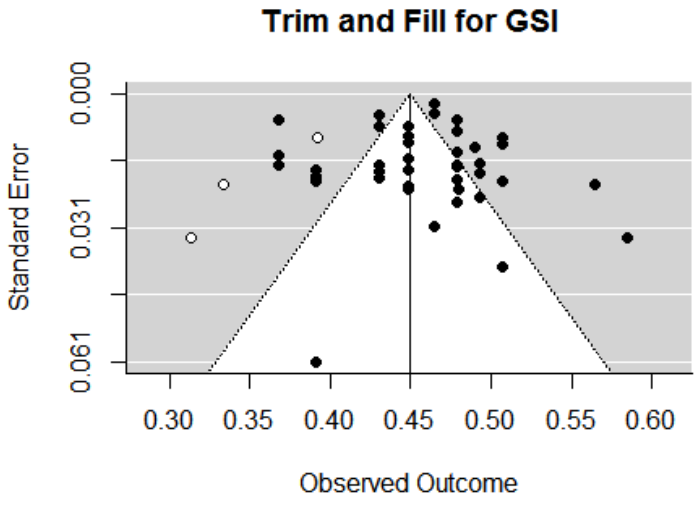

Figure 9. GSI Funnel Plot after Trim and Fill. Closed circles are original data, open circles represent filled-in data based on the trim-and-fill method.

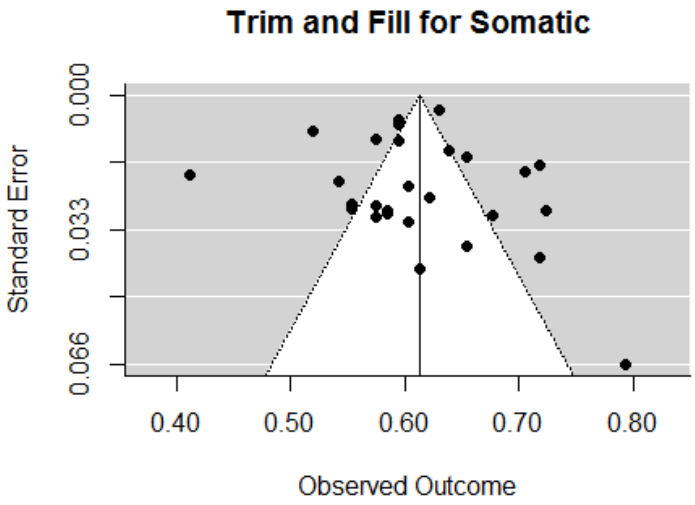

Figure 10. Somatic Funnel Plot after Trim and Fill

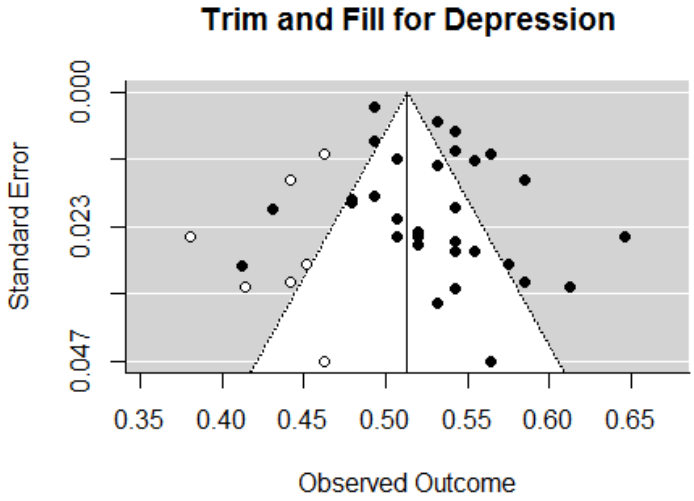

Figure 11. Depression Funnel Plot after Trim and Fill. Closed circles are original data, open circles represent filled-in data based on the trim-and-fill method.

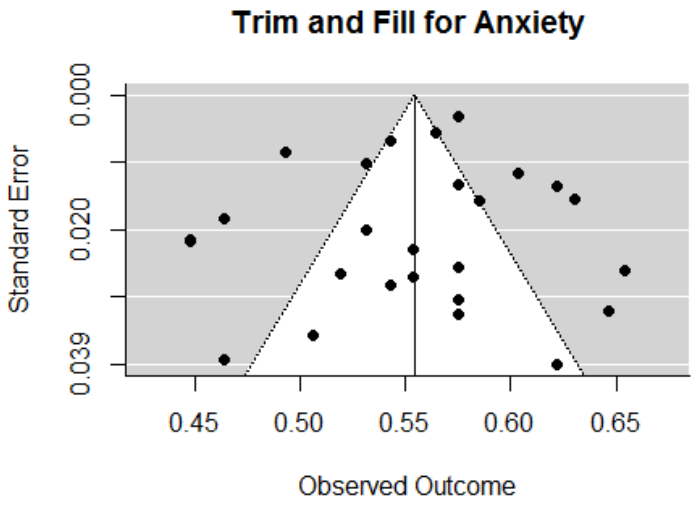

Figure 12. Anxiety Funnel Plot after Trim and Fill

The results of the moderator analysis, regression, Kendall's rank correlation, and the Egger's linear regression test indicate an absence of publication bias. Therefore, publication bias can be disregarded as a threat to the meta-analytic results.

Sensitivity Analyses

Sensitivity analysis can be defined as a technique to check the robustness of an assessment by testing the impact of changing the methods, assumptions, or values on the results (Thabane et al., 2013).

To check the robustness of the results, the analyses were repeated using the untransformed coefficient alpha. Conducting the analysis with untransformed coefficients did not show important differences compared with the results presented above. This result corresponds with López-Pina et al.'s (2015) findings that indicate similar results of using transformed and untransformed Cronbach alpha in RG studies.

Discussion

According to Hunsley and Mash (2008), when a preponderance of evidence indicates an alpha value of 0.70 to 0.79, that means the internal consistency can be considered as adequate, when the alpha is between 0.80 to 0.89, it is considered as good, and excellent when the alpha is above 0.90. According to this guideline, the GSI showed an excellent mean reliability with alpha values of 0.91, good mean reliability with alpha values of 0.83 and above for the Depression and Anxiety subscales, and the Somatic subscale showed an adequate mean reliability with an alpha value of 0.77. Hunsley and Mash (2008) point out that most authors considered 0.70 as the minimum recommended reliability. Thus, on average, the reliabilities of the BSI-18 and its three subscales were clearly above the cutoff of 0.70. However, Nunnally and Bernstein recommended a stricter criterion of 0.90 for a measure when important clinical decisions are derived from the test scores (as cited in Hunsley & Mash, 2008, p. 10). Based on this criterion, only GSI provided an appropriate reliability estimate. Even though the results of this RG meta-analysis suggest that the BSI-18 and its three subscales provide consistent information for their use with research purposes, the scores of each subscale, especially the somatic subscale, should be interpreted cautiously when these subscales are applied in clinical situations related to individual diagnosis and treatment.

Since results showed significant heterogeneity among the coefficient alpha estimates, several moderator variables were coded to determine whether they could explain the variability among the coefficient alpha estimates. Of the variables coded, population type significantly predicted the reliability estimate for the GSI. The results showed that the highest reliability estimates can be expected from the GSI with clinical samples, and lowest reliability estimate from the sample that includes both clinical and nonclinical populations. Higher reliabilities in the clinical population can be considered good news for clinical assessment, since this instrument was designed to assess these populations in particular. Also, gender significantly predicted the reliability estimate for the Somatic subscale. The results showed that the highest reliability estimates were to be expected from the Somatic subscale in samples with a higher proportion of men. Thus, researchers and clinicians should keep in mind that the reliability of the GSI measure tends to be higher in samples with a clinical sample, and the reliability of the Somatic subscale tends to be higher in samples with a higher proportion of men. Other moderators were not significant as predictors of the variability among the coefficient alpha estimates. This finding is, indeed, a positive one. It shows that regardless of the sample and measurement characteristics that were examined, the BSI-18 seems to perform in a very consistent manner. However, this finding also indicates that other moderators not considered in the model were influencing the Cronbach's alpha of the BSI-18, which future studies can test.

Practical applications

The overall scale score reliability is strong when used in its entirety (GSI). The same cannot be said for the subscales which likely include too few items, especially the Somatic scale that yielded the lower reliability estimate. It is recommended that practitioners use the GSI when conducting research or when clinically assessing participants. Even though the subscales' reliability estimates were acceptable for research, they are not recommended as the sole measure for individual use for making clinical decisions. Subscales should be used with caution if they were administered independently, because their score reliabilities did not reach the reliability estimate (0.9) that has been recommended by Nunnally and Bernstein for a measurement with a clinical purpose (Hunsley & Mash, 2008). Only the GSI would be considered appropriate for this purpose.

Limitations

Like any Meta analysis study, the main limitation of the present study is the ability to identify and include all studies that have used the BSI-18. The researcher consulted the most important database for psychology (PsychInfo). However, other databases were not considered which might provide other potential studies that can be included in the present study. Also, as mentioned above, only 29% of the studies reported Cronbach's alpha values with useful descriptive information; the lack of reliability estimates in the majority studies that used the BSI-18 was a limitation for this meta-

analysis. Given the limited number of studies reporting Cronbach alpha values, different results might have been achieved if all studies had reported these values.

References

Andjreu, Y., Galdón, M. J., Dura, E., Ferrando, M., Murgui, S., García, A., & Ibáñez, E. (2008). Psychometric properties of the Brief Symptom Inventory-18 (BSI-18) in a Spanish sample of outpatients with psychiatric disorders. *Psicothema*, *20*(4), 844-850.

Asner-Self, K. K., Schreiber, J. B., & Marotta, S. A. (2006). A cross-cultural analysis of the Brief Symptom Inventory-18. *Cultural Diversity and Ethnic Minority Psychology, 12*(2), 367-375. doi:10.1037/1099-9809.12.2.367

Card, N. A. (2012). *Applied meta- analysis for social science research*. NY: The Guilford Press.

Dalton, D. R., Aguinis, H., Dalton, C. M., Bosco, F. A., & Pierce, C. A. (2012). Revisiting the file drawer problem in meta-analysis: An assessment of published and nonpublished correlation matrices. *Personnel Psychology, 65*(2), 221-249. doi:10.1111/j.1744-6570.2012.01243.x

Devellis, R. F. (2012). *Scale development theory and applications* (3rd ed.). Thousand Oaks, Calif: SAGE Publications.

Dieckmann, N. F., Malle, B. F., & Bodner, T. E. (2009). An empirical assessment of meta-analytic practice. *Review of General Psychology, 13*(2), 101-115. doi:10.1037/a0015107

Durá, E., Andreu, Y., Galdón, M. J., Ferrando, M., Murgui, S., Poveda, R., & Jimenez, Y. (2006). Psychological assessment of patients with temporomandibular disorders: Confirmatory analysis of the dimensional structure of the Brief Symptoms Inventory 18. *Journal of Psychosomatic Research, 60*(4), 365-370. doi:10.1016/j.jpsychores.2005.10.013

Hedges, L. V., & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods, 6*(3), 203-217. doi:10.1037/1082-989X.6.3.203

Henchy, A. M. (2015). *Review and evaluation of reliability generalization research* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (Accession No. 3583730)

Hunsley, J., & Mash, E. J. (2008). *Guide to assessments that work*. Cary, GB: Oxford University Press, USA. Retrieved from http://www.ebrary.com

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.

López-Pina, J. A., Sánchez-Meca, J., López-López, J. A., Marín-Martínez, F., Núñez-Núñez, R. M., Rosa-Alcázar, A. I., & ... Ferrer-Requena, J. (2015). The Yale–

Brown Obsessive Compulsive Scale: A reliability generalization meta-analysis. *Assessment, 22*(5), 619-628. doi:10.1177/1073191114551954

Lukow, H. R., Godwin, E. E., Marwitz, J. H., Mills, A., Hsu, N. H., & Kreutzer, J. S. (2015). Relationship between resilience, adjustment, and psychological functioning after traumatic brain injury: A preliminary report. *Journal of Head Trauma Rehabilitation, 30*(4), 241-248. doi:10.1097/HTR.0000000000000137

Mason, E. J. (2007). Measurement issues in high stakes testing: Validity and reliability. *Journal of Applied School Psychology, 23*(2), 27. doi:10.1300/J370v23n02_03

Meijer, R. R., de Vries, R. M., & van Bruggen, V. (2011). An evaluation of the Brief Symptom Inventory–18 using item response theory: Which items are most strongly related to psychological distress? *Psychological Assessment, 23*(1), 193-202. doi:10.1037/a0021292

Merport, A., & Recklitis, C. J. (2012). Does the Brief Symptom Inventory-18 case rule apply in adult survivors of childhood cancer? Comparison with the symptom Checklist-90. *Journal of Pediatric Psychology, 37*(6), 650-659. doi:10.1093/jpepsy/jss050

Misono, S., Peterson, C. B., Meredith, L., Banks, K., Bandyopadhyay, D., Yueh, B., & Frazier, P. A. (2014). Psychosocial distress in patients presenting with voice concerns. *Journal of Voice, 28*(6), 753-761. doi:10.1016/j.jvoice.2014.02.010

Rodriguez, M. C., & Maeda, Y. (2006). Meta-analysis of coefficient alpha. *Psychological Methods, 11*(3), 306-322. doi:10.1037/1082-989X.11.3.306

Rowley, G. L. (1976). The reliability of observational measures. *American Educational Research Journal, 13*(1), 51-59. doi:10.2307/1162553

Slone, M., & Mayer, Y. (2015). Gender differences in mental health consequences of exposure to political violence among Israeli adolescents. *Children and Youth Services Review, 58*, 170-178. doi:10.1016/j.childyouth.2015.09.013

Spitzer, C., Hammer, S., Löwe, B., Grabe, H. J., Barnow, S., Rose, M., & ... Franke, G. H. (2011). Die kurzform des Brief Symptom Inventory (BSI-18): Erste Befunde zu den psychometrischen Kennwerten der deutschen Version. = The short version of the Brief Symptom Inventory (BSI-18): Preliminary psychometric properties of the German translation. *Fortschritte Der Neurologie, Psychiatrie, 79*(9), 517-523. doi:10.1055/s-0031-1281602

Sun, S., & Wang, S. (2015). The Children's Depression Inventory in worldwide child development research: A reliability generalization study. *Journal of Child and Family Studies, 24*(8), 2352-2363. doi:10.1007/s10826-014-0038-x

Thabane, Lehana, Mbuagbaw, Lawrence, Zhang, Shiyuan, Samaan, Zainab, Marcucci, Maura, Ye, Chenglin, . . . Goldsmith, Charles. (2013). A tutorial on sensitivity analyses in clinical trials: The what, why, when and how. BMC Medical Research Methodology, 13, 92.

Thompson, B. (1990). Alphamax: A program that maximizes coefficient alpha by selective item deletion. *Educational and Psychological Measurement, 50*(3), 585-589. doi:10.1177/0013164490503013

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement, 58*(1), 6-20. doi:10.1177/0013164498058001002

Vacha-Haase, T., Henson, R. K., & Caruso, J. C. (2002). Reliability generalization: Moving toward improved understanding and use of score reliability. *Educational and Psychological Measurement, 62*(4), 562-569. doi:10.1177/001316402128775012

Vacha-Haase, T., Ness, C., Nilsson, J., & Reetz, D. (1999). Practices regarding reporting of reliability coefficients: A review of three journals. *Journal of Experimental Education, 67*(4), 335.

Vacha-Haase, T., & Thompson, B. (2011). Score reliability: A retrospective look back at 12 years of reliability generalization studies. *Measurement and Evaluation in Counseling and Development, 44*(3), 159-168. doi:10.1177/0748175611409845

Wang, J., Kelly, B. C., Liu, T., Zhang, G., & Hao, W. (2013). Factorial structure of the Brief Symptom Inventory (BSI)-18 among Chinese drug users. *Drug and Alcohol Dependence, 133*(2), 368-375. doi:10.1016/j.drugalcdep.2013.06.017

Warne, R. (2008). *Reliability generalization (RG) analysis: The test is not reliable.* Distributed by ERIC Clearinghouse.

List of Included Studies

Andjreu, Y., Galdón, M. J., Dura, E., Ferrando, M., Murgui, S., García, A., & Ibáñez, E. (2008). Psychometric properties of the Brief Symptoms Inventory-18 (BSI-18) in a Spanish sample of outpatients with psychiatric disorders. *Psicothema,20*(4), 844-850.

Asner-Self, K. K., Schreiber, J. B., & Marotta, S. A. (2006). A cross-cultural analysis of the Brief Symptom Inventory-18. *Cultural Diversity and Ethnic Minority Psychology, 12*(2), 367-375. doi:10.1037/1099-9809.12.2.367

Canada, A. L., Murphy, P. E., Fitchett, G., Peterman, A. H., & Schover, L. R. (2008). A 3-factor model for the FACIT-Sp. *Psycho-Oncology, 17*(9), 908-916. doi:10.1002/pon.1307

Chambers, S. K., Zajdlewicz, L., Youlden, D. R., Holland, J. C., & Dunn, J. (2014). The validity of the distress thermometer in prostate cancer populations. *Psycho-Oncology, 23*(2), 195-203. doi:10.1002/pon.3391

Cieurzo, C. E. (2003). *Family environment, parental coping and distress, and socioeconomic status as predictors of psychological distress in chronically ill children* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3084907)

Cihlar, B. E. (2014). *The trauma recovery and empowerment model: A trauma-informed treatment program for female offenders in the community* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3608215)

Choau, S. T. (2014). *Effects of parental trauma experience on second generation Cambodian Americans*(Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3592956)

Cohen M, Gagin R, Cinamon T, Stein T, Moscovitz M, Kuten A. Translating 'distress' and screening for emotional distress in multicultural cancer patients in Israel. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation* [serial online]. May 2012;21(4):555-562. Available from: PsycINFO, Ipswich, MA. Accessed June 16, 2016.

Cohen, M., Mabjish, A. A., & Zidan, J. (2011). Comparison of Arab breast cancer survivors and healthy controls for spousal relationship, body image, and emotional distress. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation*, *20*(2), 191-198. doi:10.1007/s11136-010-9747-9

Collins, T. D. (2001). *Readiness Estimate and Deployability Index Revised for Air Force Nurses (READI-R-AFN) and READI-R-AFN Short Form {SF}: Psychometric evaluation*(Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3015929)

Durá, E., Andreu, Y., Galdón, M. J., Ferrando, M., Murgui, S., Poveda, R., & Jimenez, Y. (2006). Psychological assessment of patients with temporomandibular disorders: Confirmatory analysis of the dimensional structure of the Brief Symptoms Inventory 18. *Journal of Psychosomatic Research, 60*(4), 365-370. doi:10.1016/j.jpsychores.2005.10.013

France, S. J. (2007). *Spirituality and the embodied self: An exploration of the relationships between spirituality, sense of self, embodiment, and subjective well-being*(Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3240837)

Fonseca, A., Nazaré, B., & Canavarro, M. C. (2011). Patterns of parental emotional reactions after a pre- or postnatal diagnosis of a congenital anomaly. *Journal of Reproductive and Infant Psychology, 29*(4), 320-333. doi:10.1080/02646838.2011.634398

Galdón, M. J., Durá, E., Andreu, Y., Ferrando, M., Murgui, S., Pérez, S., & Ibañez, E. (2008). Psychometric properties of the Brief Symptom Inventory-18 in a Spanish breast cancer sample. *Journal of Psychosomatic Research, 65*(6), 533-539. doi:10.1016/j.jpsychores.2008.05.009

Ghadisha, H. (2005). *Somatization and its relationship to stress, acculturation and family expressiveness in iranian americans*(Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3150320)

Gianinazzi, M. E., Rueegg, C. S., Wengenroth, L., Bergstraesser, E., Rischewski, J., Ammann, R. A., & ... Michel, G. (2013). Adolescent survivors of childhood cancer: Are they vulnerable for psychological distress?. *Psycho-Oncology*,*22*(9), 2051-2058. doi:10.1002/pon.3249

Gum, A. M., Petkus, A., McDougal, S. J., Present, M., King-Kallimanis, B., & Schonfeld, L. (2009). Behavioral health needs and problem recognition by older adults receiving home-based aging services. *International Journal of Geriatric Psychiatry*, *24*(4), 400-408. doi:10.1002/gps.2135

Hawthorne, E. F. (2006). *Exploring the psychological sequelae of readjustment in citizen-soldiers deployed in military operations*(Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3214033)

Hjörleifsdóttir, E., Hallberg, I. R., Bolmsjö, I. A., & Gunnarsdóttir, E. D. (2006). Distress and coping in cancer patients: Feasibility of the Icelandic version of BSI 18 and the WOC-CA questionnaires. *European Journal of Cancer Care*, *15*(1), 80-89. doi:10.1111/j.1365-2354.2005.00620.x

Houghton, F., Keane, N., Murphy, N., Houghton, S., Dunne, C., Lewis, C. A., & Breslin, M. J. (2012). The Brief Symptom Inventory-18 (BSI-18): Norms for an Irish third-level college sample. *The Irish Journal of Psychology*, *33*(1), 43-62. doi:10.1080/03033910.2012.672247

Holakouee, F. (2011). *The relationship between acculturation and the expression of emotional distress in Iranian Americans* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3450084)

Hudgins, P. S. (2005). *The effect of racial socialization on racial identity, positive adjustment, and well being in African-American college students* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3176700)

Kim, M. A., & Yi, J. (2014). Life after cancer: How does public stigma increase psychological distress of childhood cancer survivors?. *International Journal of Nursing Studies*, *51*(12), 1605-1614. doi:10.1016/j.ijnurstu.2014.04.005

Kimron, L., & Cohen, M. (2012). Coping and emotional distress during acute hospitalization in older persons with earlier trauma: The case of Holocaust survivors. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation*, *21*(5), 783-794. doi:10.1007/s11136-011-9984-6

Lhewa, D. W. (2010). *Coping and distress among Tibetan survivors of torture and refugee trauma*(Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3382530)

Lim, J., Yi, J., & Zebrack, B. (2008). Acculturation, social support, and quality of life for Korean immigrant breast and gynecological cancer survivors. *Ethnicity & Health*, *13*(3), 243-260. doi:10.1080/13557850802009488

Martínez López, P., Durá Ferrandis, E., Vaillo, Y. A., Galdón Garrido, M. J., Pérez, S. M., & Guerra, E. I. (2012). Structural validity and distress screening potential of the Hospital Anxiety and Depression Scale in cancer. *International Journal of Clinical and Health Psychology*, *12*(3), 435-447.

Martínez, P., Andreu, Y., Galdón, M. J., & Ibáñez, E. (2015). Improving the diagnostic accuracy of the Distress Thermometer: A potential role for the Impact Thermometer. *Journal of Pain and Symptom Management*, *50*(1), 124-129. doi:10.1016/j.jpainsymman.2015.02.004

Misono, S., Peterson, C. B., Meredith, L., Banks, K., Bandyopadhyay, D., Yueh, B., & Frazier, P. A. (2014). Psychosocial distress in patients presenting with voice concerns. *Journal of Voice*, *28*(6), 753-761. doi:10.1016/j.jvoice.2014.02.010

Negi, N. J., & Iwamoto, D. K. (2014). Validation of the revised BSI-18 with Latino migrant day laborers. *Research on Social Work Practice*, *24*(3), 364-371. doi:10.1177/1049731513507980

Park, S. E. (2014). *Relationships between posttraumatic stress disorder, attachment, self-blame, and social support among women veterans with a history of childhood and/or military sexual trauma* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3599501)

Partiali, N. R. (2012). *Family conflict, psychological maladjustment, and the mediating roles of sociotropy and coping in Iranian American adolescents*(Doctoral

dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3453345)

Petkus, A. J., Gum, A. M., Small, B., Malcarne, V. L., Stein, M. B., & Wetherell, J. L. (2010). Evaluation of the factor structure and psychometric properties of the Brief Symptom Inventory-18 with homebound older adults. *International Journal of Geriatric Psychiatry*, *25*(6), 578-587.

Prinz, U., Nutzinger, D. O., Schulz, H., Petermann, F., Braukhaus, C., & Andreas, S. (2013). Comparative psychometric analyses of the SCL-90-R and its short versions in patients with affective disorders. *BMC Psychiatry*,*13*doi:10.1186/1471-244X-13-104

Recklitis, C. J., Parsons, S. K., Shih, M., Mertens, A., Robison, L. L., & Zeltzer, L. (2006). Factor structure of the Brief Symptom Inventory--18 in adult survivors of childhood cancer: Results from the Childhood Cancer Survivor Study. *Psychological Assessment*, *18*(1), 22-32. doi:10.1037/1040-3590.18.1.22

Shim, Y. R., & Schwartz, R. C. (2008). Degree of acculturation and adherence to Asian values as correlates of psychological distress among Korean immigrants. *Journal of Mental Health*, *17*(6), 607-617. doi:10.1080/09638230701506838

Slone, M., & Mayer, Y. (2015). Gender differences in mental health consequences of exposure to political violence among Israeli adolescents. *Children and Youth Services Review*, *58*170-178. doi:10.1016/j.childyouth.2015.09.013

Tremolada, M., Bonichini, S., Aloisio, D., Schiavo, S., Carli, M., & Pillon, M. (2013). Post-traumatic stress symptoms among mothers of children with leukemia undergoing treatment: A longitudinal study. *Psycho-Oncology*, *22*(6), 1266-1272. doi:10.1002/pon.3132

Torres, L., Miller, M. J., & Moore, K. M. (2013). Factorial invariance of the Brief Symptom Inventory-18 (BSI-18) for adults of Mexican descent across nativity status, language format, and gender. *Psychological Assessment*, *25*(1), 300-305. doi:10.1037/a0030436

Towns, S. J., Silva, M. A., & Belanger, H. G. (2015). Subjective sleep quality and postconcussion symptoms following mild traumatic brain injury. *Brain Injury*,*29*(11), 1337-1341. doi:10.3109/02699052.2015.1045030

Utsey, S. O., & Hook, J. N. (2007). Heart rate variability as a physiological moderator of the relationship between race-related stress and psychological distress in African Americans. *Cultural Diversity and Ethnic Minority Psychology*, *13*(3), 250-253. doi:10.1037/1099-9809.13.3.250

Van Dyke, C. J. (2011). *The relationship of perceived stress, general coping, and religious coping to psychological distress and adjustment in urban high-school students*(Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3438467)

Wang, C. D., & Mallinckrodt, B. (2006). Acculturation, attachment, and psychosocial adjustment of Chinese/Taiwanese international students. *Journal of Counseling Psychology*, *53*(4), 422-433. doi:10.1037/0022-0167.53.4.422

Wenninger, K., Helmes, A., Bengel, J., Lauten, M., Völkel, S., & Niemeyer, C. M. (2013). Coping in long-term survivors of childhood cancer: Relations to psychological distress. *Psycho-Oncology*, *22*(4), 854-861. doi:10.1002/pon.3073

Wetherell, J. L., Birchler, G. D., Ramsdell, J., & Unützer, J. (2007). Screening for generalized anxiety disorder in geriatric primary care patients. *International Journal of Geriatric Psychiatry*, *22*(2), 115-123. doi:10.1002/gps.1701

White, L. A. (2004). *Effects of psychological distress on employment among low-income women*(Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3137269)

Young, A. C. (2015). *Effects of the interaction of religion and internalized homonegativity on psychological well-being* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3639853)

Zabora, J., BrintzenhofeSzoc, K., Jacobsen, P., Curbow, B., Piantadosi, S., Hooker, C., & ... Derogatis, L. (2001). A new psychosocial screening instrument for use with cancer patients. *Psychosomatics: Journal of Consultation and Liaison Psychiatry*, *42*(3), 241-246. doi:10.1176/appi.psy.42.3.241

Appendixes

Appendix A

RG of the BSI-18 Coding Book

Note: 0 = N/A, Not Reported, or No for all coding categories

Report Identification
ID code # (start with 01)

Research study identification (Citation)
Author(s) (author's names – last name, first name)
Year of publication
Publication type
1. Journal
2. Conference proceedings (paper)
3. Organization (report)
4. Dissertation or Thesis
5. Other

Research quality
1. Not published
2. Published study

Sample characteristics
Sample size N (Value)
The average age of the sample (Value)
Gender
Female percent (Value)
Population type
1. Non- clinical
2. Clinical
3. Both clinical and non-clinical

Instrument characteristics
The language of the inventory
1. English

2. Non English
3. Two languages

The Cronbach's alpha values

The Cronbach's alpha value of GSI or the total score (Value)
The Cronbach's alpha value of Somatic subscale (Value)
The Cronbach's alpha value of Depression subscale (Value)
The Cronbach's alpha value of Anxiety subscale (Value)

Appendix B

RG of the BSI-18 Coding Form

Note: 0 = N/A, Not Reported, or No for all coding categories

Report Identification
- ID code #: _____

Research Study Identification (Citation)
- Author(s) _____
- Year of Publication _____
- Publication Type _____
- Research Quality _____

Sample characteristics
- Sample size N (Value) _____
- The average age of the sample (Value) _____
- Gender
  - Female percent (Value) _____
- Population type_____

Instrument characteristics
- The language of the inventory_____
- The Cronbach's alpha values
  - The Cronbach's alpha of GSI or the total score (Value) _____
  - The Cronbach's alpha of Somatic subscale (Value) _____
  - The Cronbach's alpha of Depression subscale (Value) _____
  - The Cronbach's alpha of Anxiety subscale (Value) _____